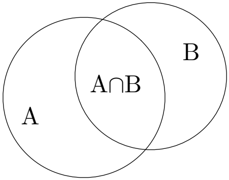
Gaussian processes: The Bayes Rule

Manel Martínez-Ramón

ECE, UNM

October, 2018

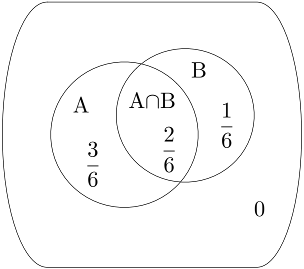
Assume a universe of possible events Ω and two subsets $A, B \in \Omega$



- $P(A|B) =$ probability of A given that B ocurred.
- Then B is our new universe.
- Bayes rule: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

The intuitive interpretation is that since B occurred, then the rest of the universe disappears, as its probability is 0. The conditional probability is the fraction of areas between $A \cap B$ and B.

Bayes rule: an example



- $P(A) = \frac{3}{6} + \frac{2}{6} = \frac{5}{6}$
- $P(B) = \frac{2}{6} + \frac{1}{6} = \frac{3}{6}$
- $P(A \cup B) =$
  $P(A) + P(B) - P(A \cap B) = 1$
  Hence,
  $P(A \cap B) = \frac{5}{6} + \frac{3}{6} - 1 = \frac{2}{6}$

Then the probability of A given that B has ocurred is
$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}$

Corollary

- We can do inference over an event $A|B$ known the distribution of $B|A$ and the marginal distribution B.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow \begin{cases} P(A \cap B) = P(A|B)P(B) \\ P(B \cap A) = P(B|A)P(A) \end{cases}$$

- Since $A \cap B = B \cap A$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes Rule on several conditional variables

- Now define the r.v. $\mathbf{w}$ and $y \cap x = y, x$ and compute $p(\mathbf{w}|y, x)$

$$p(\mathbf{w}|y, \mathbf{x}) = \frac{p(y, \mathbf{x}|\mathbf{w})p(\mathbf{w})}{p(y, \mathbf{x})}$$

$$p(y, \mathbf{x}) = p(y|\mathbf{x})p(\mathbf{x}), \qquad p(y, \mathbf{x}|\mathbf{w}) = p(y|\mathbf{x}, \mathbf{w})p(\mathbf{x})$$

- Then

$$p(\mathbf{w}|y, \mathbf{x}) = \frac{p(y|\mathbf{x}, \mathbf{w})p(\mathbf{x})p(\mathbf{w})}{p(y|\mathbf{x})p(\mathbf{x})} = \frac{p(y|\mathbf{x}, \mathbf{w})p(\mathbf{w})}{p(y|\mathbf{x})}$$

- Finally:

$$p(\mathbf{w}|y, \mathbf{x}) \propto p(y|\mathbf{x}, \mathbf{w})p(\mathbf{w})$$

Conclusion

- If w is a set of parameters of the estimator

$$y = \mathbf{w}^\top \mathbf{x} + \varepsilon$$

  we call $p(\mathbf{w})$ the *prior* probability, and then $p(\mathbf{w}|y, \mathbf{x})$ is the *posterior* probability.

- Similarly, we can compute the output *likelihood* using the expressions of slide 5 as

$$p(y|\mathbf{x}, \mathbf{w}) = \frac{p(y, \mathbf{x}|\mathbf{w})}{p(\mathbf{x})}$$